

ABSTRACT OF THE DISCLOSURE

A system to load data in a data warehouse includes reception of a plurality of records, determination, for each of the plurality of records, of values representing differences between a record and each other of the plurality of records, identification of at least two of the plurality records as duplicates based on a determined value representing a difference between the two records, and storage of the two records in the data warehouse in association with a same identifier. Determination of the values may include determination, for each of a first plurality of data fields of the record, of a first value representing a difference between data specified in the data field and data specified in a respective one of a second plurality of data fields of one of the other of the plurality of records, determination, for each of the second plurality of data fields, of a second value representing a difference between data specified in the data field and data specified in a respective one of the first plurality of data fields, and determination of a third value representing a difference between the record and the one of the other of the plurality of records based on the determined first and second values.